The Predictive Validity of Language Assessment in a Pre-University Programme in two Colleges in Oman: A Mixed-Method Analysis

Fatma Al Hajr[1]

Abstract

This study investigates the predictive validity of English language assessment in two Colleges of Applied Sciences in Oman. It also looks at the language requirements of higher education academic programmes through analysing the assessment instruments used. It was conducted over two academic semesters, in the first semester the grades of 184 Foundation Programme students in English language courses were obtained, in the second semester the grades of 163 First Year students in academic courses were obtained. The predictive validity of English language assessment with regards to academic achievement was found to be $r=0.3$, $p < 0.01$, and the strength of the predictive validity significantly differed among specialisation and self-evaluation groups. Test papers and assessment tasks used in three academic programmes (Information Technology, International Business Administration and Communication Studies) were also analysed to explain the findings on English language assessment predictive validity. The implications of the findings on national and international higher education are discussed and recommendations are presented.

Proficiency in English language and how it is measured have become a central issue in higher education research as the English language is increasingly used as a medium of instruction and a criterion for admission to higher education worldwide.

---

[1] Ministry of Higher Education, Oman. E-mail: fatimaalhajri@gmail.com

In a review of a number of articles about language policies in Asian higher education, Ross (2008, p. 8) states that "a commonly accepted assumption is that a foreign language learned in the context of formal schooling yields suitable subject matter for making high-stakes inferences about qualifications for admissions or employment"; he explains that there is a growing use of test scores in determining access to higher education, and that proficiency in the English language has also become a dominant criterion for success in the labour market. Following this phenomenon, proficiency in the English language has been considered a criterion to access most higher education programmes in Oman, and the English language assessment plays a critical role in admission to higher education. However, there is always a question about how predictive student scores in English language assessment are of student success in future academic study. This paper investigates the predictive validity of student scores in English language assessment in terms of following academic achievement in two Colleges in Oman, and it analyses the English language requirements of three academic courses from the Information Technology (IT), Communication Studies (CS) and International Business Adminstration (IBA) departments.

The Meaning of Assessment Validity and Predictive Validity

Test validity has been mainly viewed as five separate validities (i.e., face, content, predictive, concurrent, and construct) that represent distinct psychometric characteristics of a test. Sometimes these validities are grouped into internal, external and construct validities. The internal validity consists of face validity and content validity, whereas the external or criterion validity (Martuza, 1977) consists of concurrent validity and predictive validity. Hughes (2003, pp.22-28) explains the meaning of each type, saying that face validity signifies that an assessment looks suitable for its purposes, content validity means that an assessment is representative of the skills and content which is supposed to measure, concurrent validity is established when an assessment correlates well with another test that similarly assesses the same constructs and is undertaken at about the same time, predictive validity means the extent to which an assessment predicts future performance of assessed participants, and construct validity indicates that an assessment instrument measures the skills and abilities (i.e., constructs) that it is supposed to be measuring.

Content validity is concerned with identifying the relationship between test tasks and specific learned content while construct validity attempts to make the connection between test tasks and theoretical constructs of language proficiency regardless of learned materials.

This view sees reliability as a distinct quality from validity but both are necessary for a good test. A test's reliability is shown if similar scores are obtained when the same test is administered to two groups equal in the abilities or administered to one group at different times (Hughes, 2003). Harrison says "the *reliability* of a test is its consistency" (1983, p.10, italics in original).

This view was challenged by Messick's seminal article (1989) in which he not only redefined validity as a unitary concept that involved multiple facets, but also argued that the consequences of a test should be included as an aspect of validity. He affirmed that the consequences of a test constituted an inherent facet of any evaluative judgement of the "*adequacy* and *appropriateness* of *interpretations* and *actions* based on test scores" (Messick, 1995, p.5, italics in original). Validity was defined as "a unified though faceted concept", and validation as a "scientific enquiry into score meaning" (Messick, 1989, p.6). Test validity in this unitary understanding still consists of the former validities, but they are seen as aspects not independent entities, and they are encompassed by the overarching construct validity which links evidence from all other aspects, including the novel consequential aspect, to constitute one comprehensive concept. Bachman (2004) clarifies the premises of validity in Messick's view saying that (a) validity indicates the quality of the interpretation not scores, (b) validity is a matter of a degree and is not static, (c) validity is specific to a particular use, and (d) validity involves a comprehensive evaluative judgment. In this view, test validation is presented as the process of collecting information that supports the appropriateness and correctness of the interpretations of the test scores (Messick, 1989; Bachman & Palmer, 1996; McNamara, 1996). Thus, when a test is used for a purpose that it was not designed to fulfill, it becomes invalid (Baker, 1989). The process of validating an assessment instrument should examine the evidence provided for the claims made by its developers about the scores' interpretations.

Studies on the Predictive Validity of IELTS, TOEFL and In-house Tests

In spite of the widespread theoretical acceptance of the unitary view of validity that involves several 'aspects', studies on the predictive validity of language assessment are still carried out for their own merits (i.e., estimating future performance by correlating results on two different assessment instruments separated by a specific time difference).

Twenty-five years ago, Graham (1987) described the results obtained from predictive validity studies on language tests as inconsistent, and the same conclusion can be drawn today from the following selective summary in Tables 1-3 which is divided into studies about the predictive validity of internationally used tests as gatekeepers to higher education institutions (i.e., IELTS, TOEFL) and in-house language tests.

Table 1. Some Studies on Predictive Validity of IELTS

Study	Country	Number of Participants	Type of Correlation	Correlation strength
Elder (1993)	Australia	32 International Students	IELTS & Administrator Ratings	0.5*
Cotton & Conrow (1998)	Australia	33 Undergraduate & Postgraduate Students	IELTS & GPA	-0.24*
			ILTS & Staff Ratings	0. 15*
			IELTS & Student Self-assessment	-0.28*
Huong (2001)	Australia	320 Vietnamese Post- & Undergraduate Students	IELTS & GPA	0.30*
Kerstjen & Nery (2000)	Australia	113 International Students	IELTS & GPA	Non-Significant
Feast (2002)	Australia	101 International Students	IELTS & GPA	0.39*
Woodrow (2006)	Australia	62 Students 15 Teachers in Faculty of Education	IELTS & Teacher Evaluations	0.40*
Breeze & Miller (2008)	Spain	289 Undergraduate Spanish Students	IELTS & GPA (Humanities)	0.34*
			(Law)	0.28 **
			(Medicine)	0.25*
Yen & Kuzma (2009)	Britain	61 Chinese Students (Bussiness)	IELTS & GPA	0.46**

* $p < 0.05$, ** $p < 0.01$

Table 2. Some Studies on Predictive Validity of TOEFL

Study	Country	Number of Participants	Type of Correlation	Correlation
Vinke & Jochems (1993)	Netherlands	90 Indonesian Students (Engineering)	TOEFL & GPA	TOEFL< 450 = 0.09**
				TOEFL > 450 = 0.5**
Cho & Bridgeman (2012)	USA	2594 Graduate & Undergraduate Students	TOEFL & GPA	Graduate Students = 0.16*
				Undergraduates = 0.18*
Al-Musawi & Al-Ansari (1999)	Bahrain	86 Undergraduate Students (English Language Studies)	TOFEL & GPA/ENGPA***	GPA = 0.50**
				ENGPA = 0.70**
Maleki & Zangani (2007)	Iran	50 Undergraduate Students (English language studies)	TOFEL & GPA	0.48*

*p < 0.05, **p < 0.01, *** Students' GPA in English Language Major

Table 3. Some Studies on Predictive Validity of In-house Language Tests

Study	Country	Number of Participants	Type of Correlation	Correlation
Davies (1990)	UK	310 International Students	ELBA & GPA	0.30**
Lynch (2000)	UK	475 International Students	TEAM1 & GPA	0.32*
		291 International Students	TEAM 2 & GPA	0.28*
Jochems, et al. (1996)	Netherland	170 International Students	Dutch Exam & GPA	0.36**

 *p < 0.05, **p < 0.01

Factors Affecting the Predictive Validity of Language Assessment Specialisations

A variance in the strength of the predictive validity values of language assessment across specialisations has been reported by several studies of different specialitations.

Jochems et al. (1996) found that the value of the predictive validity varied from r = 0.32 to r = 0.46 in Computer Sciences and Engineering majors. Their study looked at the correlations between Dutch language proficiency as a second language (Dutch was the medium of study) and academic achievement. Lynch (2000) found that there was some difference in the correlation coefficient between the English language test used at the University of Edinburgh and students' average scores in the academic courses across the students' different fields of study. For example, the correlation coefficients in the Arts and Veterinary Medicine were non-significant, whereas, the coefficients in Social Sciences, Law, Science and Engineering were r = 0.23, r = 0.32 and r = 0.24 respectively. Similarly, Huong (2001) claimed that the correlation between language proficiency and academic achievement in the linguistically demanding disciplines (e.g., TESOL) was stronger than it was in the less linguistically demanding disciplines (e.g., Engineering). Woodrow (2006) reported the correlation coefficient between the students' bands in IELTS and their GPA in TESOL courses to be r = 0.4, p < 0.01, *n* = 62. In the English language domain, Cope (2011) reported that the value of the correlation varied between different disciplines when he studied the predictive validity of three types of English language entry programmes.

Self-Evaluation of Language Skills

Very few studies on predictive validity have investigated the possible contribution of the students' self-evaluations to the strength of the predictive validity of presessional language assessment (Powers, Kim, & Weng, 2008; Xu, 1991), the second of these produced interesting results. Xu (1991) investigated the correlation between students' self-evaluations of their language proficiency and self-reported academic difficulties, and the correlation between TOEFL scores and self-reported academic difficulties. Xu found that the students' self-evaluation was a better predictor of the perceived academic difficulties than were their TOEFL scores. Though Xu's focus was on perceived academic difficulties, his findings draw attention to the role of self-evaluation in understanding possible future academic difficulties.

Methodology

Given that validation is a "scientific inquiry into score meaning" (Messick, 1989, p.6), this paper investigates the score interpretation that assumes a positive correlation between student scores in English language assessment and their scores in academic courses taught in English.

This paper reports on the findings of a study conducted in partial fulfillment of the requirements of the degree of a Doctor of Philosophy at The University of Edinburgh (Author, 2013; Author, 2014).

We need predictive validity studies as one type of evidence towards verifying the claims and inferences made using test scores (Bachman, 2004; Bachman & Palmer 1996; Kane, 2011; Messick, 1995; Weir, 2005). In this study, students' grades in the Foundation Programme (FP) assessment in two Colleges of Applied Sciences (CAS) are correlated with their grades in the academic courses of the first semester of their First Year (FY), which actually is in the following academic semester. Students started the FP in February 2011 and the FY in September 2011. The predictive validity of FP assessment is investigated in groups of college, gender, self-evaluation, and specialisation. Linguistic requirements of three academic courses (IT, CS, and IBA) are analysed using more than 20 documents including, course learning outcomes, syllabi and test papers.

Table 4. Assessment Instruments in the Foundation Programme Courses (Author, 2013, p.205)

Course	Assessment Instruments	% of Course Total	% of Foundation Programme Total
General English	Midterm Test	40%	50%
	Final Test	60%	
Academic English	Presentation	50%	50%
	Report	50%	

The FP is a pre-sessional programme that consists of two hours of mathematics and/or computer skills courses in each semester. The English language programme is divided into two major courses, the General English skills (GES) and Academic English Skills (AES)

AES assessment includes continuous assessment (i.e., a report and presentation) as shown in the Table 4. GES assessment included tests which were centrally developed, though individual teachers from different colleges participated in the process of writing, reviewing and rewriting these tests. The teachers participated in standardization and moderation training sessions prior to marking the writing component of GES tests.

Though these tests were constructed and reviewed following rigid procedures, they were not trailed before use and their reliability is unknown. Similarly, AES assessment utilized rating scales to evaluate student performances in the report and presentation, however, no sessions in standardizing the implementation of the rating scales in AES were offered to teachers. For the purposes of this study, the term (FP) will be used to refer to the English component only (i.e., GES and AES). Student scores in Mathematics and Computer Skills were not included in this predictive validity study of FP assessment.

Research Questions

This study investigated the predictive validity of FP assessment by correlating students' scores in FP assessment and their scores in three FY academic courses. It also investigated the linguistic demands of these three FY academic courses. It investigated the questions below:

1. Did student performance in English language assessment in FP correlate positively with their performance in academic courses assessment?
2. Did the strength of correlation between the language proficiency and academic achievement differ significantly when students' scores in GES assessment or AES assessment only were used, instead of the overall scores (i.e., FP) in both?
3. Did the groupings by college, gender, self-evaluation and specialisations show significant differences in the correlations between language proficiency and academic achievement?
4. What were the linguistic demands of the academic courses in the FY as deduced from analysing the learning-outcomes and tests of these courses? How can these language demands explain the findings on FP assessment predictive validity?

Operational Definitions of 'Proficiency' and 'Achievement'

Before investigating the relationship between the students' language proficiency and their academic achievement, it is crucial to explain how the concepts 'language proficiency' and 'academic achievement' were operationalised. Students' English language proficiency was represented by their average grades on the two FP English language courses (i.e., AES and GES). Likewise, the students' achievement in academic courses was represented by their average grades on the FY academic courses in the first semester.

Scores on courses unrelated to the specialisations or on those taught in Arabic (e.g., Islamic Culture or Omani Economic History courses) were not included in calculating the students' average scores on the academic courses.

Another point to clarify is how the Grade Point Average (GPA), used in CAS to report students' achievement, was employed in this study. GPA stands for "the Grade Point Average of the numeric value of the entire results that the student has passed or failed in that semester" (CAS, 2010, p. 4). To calculate the GPA, student scores were transformed from numeric grades to grade points ranging from 0 to 4 using the scale in Table 5, which was also the standard scale for calculating GPA in CAS. The crude GPA form of the FY was deemed to be unsuitable for this study as it included the average results of all of the courses taken in a specific semester. This study investigated only the English language medium courses that were related in content to the students' academic specialisations. Therefore, only the grade points of the academic courses that were taught in English and related in content to the students' academic study were included in the GPA used to represent academic achievement.

Table 5. Conversion Table for Scores Used in CAS*

Numeric Grade	<50	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-100
Grade point	0	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.3	3.7	4.0
Letter Grade	F	D	D+	C-	C	C+	B-	B	B+	A-	A

*. from the Registration Office at Sur CAS, personal communication, February 14, 2012

One complication encountered was that the students' scores in the academic courses were only available in a grade point system, while their scores in the FP assessment were available in a numeric system. To overcome having the grades in two different forms, scores in the FP were converted to grade points utilising the scale used in CAS and shown in Table 5. For example, if a student's score in FP was between 80 and 84, this score was converted to a grade point of 3.0.

Data Analysis

Statistical Analyses Used with the Student Scores

This study investigates the correlation between students' English language proficiency on the FP (measured by their scores in the two English language courses assessments) and their academic achievement in the FY (measured by their average scores in the first semester of the FY assessment). For more discussion on correlational analysis in applied linguistics, see Dörnyei (2007, pp.223-241). It also focuses upon whether the strength of the correlation was affected by the different groups of students. Two types of statistical analyses were applied, namely correlational analysis using Spearman's rho and the difference in means analysis using Mann-Whitney U test and the Kruskal Wallis Test. The latter tests were used to identify significant differences between student scores in different groups when the predictive validity varied amongst the groups. Non-parametric tests were used because the distribution of the scores was negatively skewed and the sizes of the group samples were not equal (Pallant, 2007).

Thematic Content Analysis of the Course Documents

In this study, thematic (ethnographic) content analysis, which focuses on "what is said rather than on how it is said" (Bryman, 2008, p.412) was used to analyze the course documents. The parameters followed in coding and analysing are explained in other papers on this research project (Author, 2013; Author, 2014).

Participants

The study was conducted over two academic semesters, in the first semester the grades of 184 FP students were obtained, but in the second semester the grades of only 163 FY students were obtained.

This decrease in participation was caused by (1) some students' inability to pass Foundation Programme assessment, or (2) taking Arabic medium courses only in the First Year. The participants were recruited from two CAS colleges, namely: Sur and Rustaq.

Results

FP Assessment Predictive Validity

Students' grades in the FP English language courses and their average grades in the FY academic courses were tested for normality of distribution using Kolmogorov-Smirnov, Shapiro-Wilk tests and histograms. The results showed that the students' scores were all negatively skewed (see Table 6). For this reason, only non-parametric statistical tests were used to explore the dataset (see Pallant (2007) for more discussion on non-parametric tests).

Table 6. Descriptive Statistics and Skewness of Student Scores in FP and FY Academic Courses assessment

Courses		N	Minimum	Maximum	Mean	Std. Deviation	Skewness
FP Assessment	AES assessment	143	1.70	4.00	3.22	.48	-.36
	GES assessment	143	.00	3.70	2.23	.62	-.72
	(AES + GES)	143	1.0	4.0	2.77	.47	-.62
FY Academic Courses Assessment		164	.50	3.90	2.71	.68	-1.08

FY academic courses are English language mediated and are considered introductory courses in the IT, IBA and CS academic programmes. The results showed a highly significant, but weak correlation between the two variables, rho=0.31, p < 0.01 (see Table 7). In addition, the difference in the predictive validity of each of the FP courses (i.e., GES and AES) was explored. The students' grades in the GES assessment correlated weakly with their average grades in the academic courses, rho =0.37, p < 0.01.

However, the correlation between the students' grades in the AES assessment and in the academic courses assessment was weaker, rho=0.27, p < 0.01. In other words, the students' grades in the FP assessment were generally a weak predictor of their grades in the academic courses.

Table 7. Correlations between Students Grades in Academic Courses, Foundation Programme assessment, General English Skills Test and Academic English Skills Assessment

Courses	Academic Courses (N=163)	FP (GES +AES) (N=163)	GES (N=163)	AES (N=163)
Academic Courses	1.000	.311**	.367**	.272**
FP	.311**	1.000	.806**	.824**
GES	.367**	.806**	1.000	.474**
AES	.272**	.824**	.474**	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

Student scores in AES assessment were much higher than their scores in GES assessment as Table 6 shows; however, the predictive power of GES assessment is higher than AES assessment. A possible explanation for the higher correlation between student scores in GES assessment and Academic Courses assessment is the type of assessment instrument used (i.e., tests). It is worth remembering at this point that the GES assessment like FY assessment consisted of standardised tests while the AES assessment consisted of performance assessment tasks.

Comparing the Predictive Validity of FP across the Groups

Differences between College Groups

The predictive validity of the English language assessment in FP was stronger for the participants from Sur College than it was for those from Rustaq College. The table below shows that Spearman coefficients for the students' grades in the FP and FY assessment were rho = 0.46, p = 0.002 for Sur College (n = 44); and rho = 0.16, p = 0.088 for Rustaq College (n = 199).

It is worth noting that the correlation between the scores in the FP and FY assessment was found to be non-significant for the Rustaq College group. This difference is further explained in the following sections.

Table 8. Correlation between Scores in FP and FY Assessment by Colleges

College	Correlation	Sig.	*N =163*
Rustaq	.16	.088	199
Sur	.46**	.002	44

**. Correlation is significant at the 0.01 level (2-tailed).
Foundation Programme (FP), First Year (FY)

Differences between Gender Groups

The correlations between the students' scores in the FP assessment and their grades in the FY academic courses assessment were not very different between the gender groups. The Spearman coefficient for the male group was rho = 0.30 and for the female group rho = 0.32.

Table 9. Correlation between Scores in FP and FY assessment by Gender

Gender	Correlation	Sig.	*N=163*
Male	.30*	.07	61
Female	.32**	.000	102

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
Foundation Programme (FP), First Year (FY)

Differences among Self-evaluation Groups

The students had been asked to self-evaluate their language proficiency using the descriptors: weak, average, good, very good and excellent. The Spearman correlation between students grades in the FP assessment and their grades in FY academic courses assessment ranged from rho = 0.17 for the *average* group to rho = 0.88 for the *excellent* Group (see Table 9).

This means that the higher the students evaluated their language proficiency the stronger the predictive validity coefficient of FP assessment became, and consequently the more their performance in the academic courses assessment became predictable by their performance in the FP assessment.

Table 10. Correlations between scores in FP and FY assessment by to Self-Evaluation Groups

Self-Evaluation	Correlation	Sig.	$N = 163$
Average	.17	.59	15
Good	.25*	.02	85
V. Good	.39**	.005	51
Excellent	.88**	.009	12

*. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Foundation Programme (FP), First Year (FY)

Differences among Specialisation Groups

Interestingly, the strength of the predictive validity of the FP assessment varied depending on the students' specialisations. Table 10 shows that the students' grades in IBA and IT courses were less well predicted by their grades in the FP assessment than were their grades in CS and English language (education) courses. The predictive validity of FP assessment in the specialisation groups ranged from rho = 0.18, p = 0.12 for the IBA group to rho = 0.64, p = 0.002 for the CS group.

Table 11. Correlations between Scores in the FP and FY Assessment by Specialisations

Specialisation	Correlation	Sig.	$N = 163$
Information Technology (IT)	.41*	.008	41
Communication Studies (CS)	.64**	.002	21
International Business Administration (IBA)	.18	.12	78
English Language-Education	.57**	.005	23

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
Foundation Programme (FP), First Year (FY)

The difference in the predictive validity between the two Colleges could be explained by the type of specialisations taught in each of the colleges and the size of student samples represented by each specialisation in this study (see Table 11). The participants from Sur College were specialised in either IT or CS; and the participants from Rustaq College were specialised in either IT, IBA or English language (Education major).

The fact that most of the Rustaq College participants were IBA students (66.93% of the sample), and that the predictive validity of FP asessment for the IBA group was non-significant, could very well explain the non-significant result obtained for the predictive validity of the FP assessment in the Rustaq group.

Table 12. The FP assessment Predictive Validity by College and Specialisation

College	Specialisation	Correlation	Sig.	n
Rustaq	IT	.27	.27	18
	IBA	.11	.31	78
	English Language-Education	.66**	.001	23
Sur	IT	.14	.52	23
	CS	.73**	.000	21

**. Correlation is significant at the 0.01 level (2-tailed).

The findings of this study suggest that the predictive validity of FP assessment is weak. However, the strength of the predictive validity varies depending on the student specialisations and self-evaluations; the predictive validity of FP assessment is stronger for CS students and the students who evaluate their language skills as higher.

Language Requirements of the Academic Courses in the First Year

This section attempts to understand the link between the language skills focused upon by the FP assessment and the ones required by the FY academic courses assessment. This is done by analysing the relevant FY course documents (i.e., syllabus, test papers, and course specifications). This may help to clarify, add meaning to and provide a counterweight to the numerical findings on FP assessment predictive validity reported previously. Therefore, this section first explores the language demands implied in the FY academic course specifications, and then investigates the linguistic nature of the written output required in the coursework and final tests for these courses.

Comparison of the FP English Syllabus and the FY Academic Courses Syllabi

In order to understand the language focus of the academic courses, the syllabi of the introductory courses of the Information Technology (IT), International Business Administration (IBA) and Communication Studies (CS) were analysed to identify the learning outcomes that seemed to demand linguistic skills. These learning outcomes were compared with those of the FP. Table 12 displays the learning outcomes of these courses. Those that seem to require complex English language output in the academic courses syllabi, and the FP outcomes that seem to match the academic courses' linguistic demands, are highlighted. One learning outcome is highlighted in the IT course, three in the IBA course, three in the CS course and three in the FP course.

The initial comparison of the highlighted learning outcomes in the FP and FY courses suggests that most of the language skills drawn upon by the academic courses' outcomes were covered by the FP outcomes. For example, in the FP, students were expected to master discussing issues in written and oral forms (see points 3 and 6, row 1 in Table 13). These two learning outcomes seem to correspond with the linguistic demand of discussing or explaining concepts entailed in the learning outcomes list of the IT and IBA courses. Similarly, the FP learning outcome of being able to read around 1000 word essays (see point 2, row 1) could presumably equip the students with the skills needed to understand or identify certain concepts from reading passages as required by all of the academic courses' learning outcomes.

Table 13. The Learning Outcomes of the FP English, IT, IBA and CS Courses

Course	FP English (Foundation English Course Specification, 2010, p. 18 & p.19)	IT (Fundamentals of Information Technology, 2008, p.1)	Communications (An Introduction to Personal Communication: Student handbook, 2008, p.2)	IBA (Bachelor of International Business Administrations, 2008, p.23)
Objectives	• Read an extensive text of around 1000 words broadly relevant to an area of study and respond to questions that require analytical skills, e.g. prediction, deduction, inference. • Produce a written report of a minimum of 500 words showing evidence of research, note taking, review and revision of work, paraphrasing, summarising, use of quotations and use of references. • Take notes on peer presentations, sufficient to enable the student to re-construct the	• An introductory understanding of computer systems, their components, and their interactions • Competence with application software, in particular word processing, spread sheets and graphics programs • An understanding of both why good ergonomic practices are important, and how to apply them in a personal context • An introductory understanding of the development of the Internet, the World Wide Web, and multimedia; their interactions and common uses/applications, in particular e-commerce. • The ability to discuss the impact of computer technology on society	• Demonstrate an understanding of the basic concepts involved in the communication process. • Identify the reasons for communication breakdown. • Demonstrate a basic understanding of non-verbal communication cues. • Demonstrate the skills necessary to give a competent oral presentation. • Identify and practise the basic factors involved in effective group work. • Demonstrate an understanding of the cultural factors which have an effect on communication.	• Identify the factors that influence the contemporary business environment; • Discuss the challenges of business, with a focus on the Omani context; • Recognise issues and concerns (e.g., accounting, marketing, finance) related to current business scenario; • Explain the relationship of business to socio-economic conditions; and • Demonstrate an interest to manage an entrepreneurial undertaking.

	main points of the presentation. • Take notes on longer talks/mini-lectures (10-15 minutes) • Prepare and deliver a talk of at least 5 minutes. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions.	• An understanding of study paths and career opportunities in information technology • A broad understanding of ethical concepts related to computing.		

Nonetheless, a comparison of the courses' learning outcomes provides brief information about what language skills were required in the FY academic courses and what was offered by the FP English language courses. Therefore, the assessment used in the IT, IBA and CS introductory courses in the first semester of FY are reviewed in the next section to obtain a deeper understanding of how and what English language skills were required in the FY academic study.

Investigating Assessment of the Academic Courses

The Course Work

The types of assessment tasks given to the students in the first semester of their FY study were analysed in terms of the apparent language requirements.

Assessment in the IT course can be divided into two parts: course work which, as stated in the course specifications, evaluated the practical skills imparted during the course, and a final test which evaluated the students' understanding of the main theoretical concepts introduced in the course. In the IT course, a graphics assignment, lab work and lab exam were used to evaluate certain IT skills; they used basic language in the instructions and required very limited language responses (see Table 14). Similarly, the IBA coursework included a series of multiple-choice quizzes and individual e-learning activities which seemed to require moderate language use, and a final test to evaluate students' understanding of the focal theoretical concepts and issues. The CS coursework, on the other hand, used assessment tasks that seemed to demand a good mastery level of the English language, such as delivering an informative talk and writing a 1000 word essay. It also included a final test to evaluate the students' grasp of the main concepts introduced in the course. The types and weightings of the instruments used in the coursework part of assessment in the academic courses are displayed in the table below.

Table 14. Assessment Instruments in FY Academic Courses

Course	Assessment	Weightings
IT	Graphics assignment	15
	Completion of lab work	15
	Lab exam	35
	Final test	35
	Total	**100**
IBA	Quiz 1	10
	Quiz 2	10
	Group assignment, e-learning activities	20
	Final test	40
	Total	**100**
Communication Studies	3-Minute informative talk	15
	1000 word essay	25
	5-Minute persuasive presentation	20
	Final test	40
	Total	**100**

The Final Test

The IBA and CS final tests constituted 40% of the total course mark, and the IT final tests constituted 35% of the final mark, but the test papers themselves were designed to be marked out of 50, 100, or 80 respectively; this was then converted to the mentioned percentage of the total course weightings. The IT and IBA test tasks both utilised the multiple choice and true/false format in the first section of the tests. In the second section, they both used short question format that required defining concepts or mentioning elements of a concept, and could be answered by mere memorization and did not seem to involve much original language use (see Table 14). The long answer questions used in the third section of both tests seemed not very different from the short ones in terms of the language output they required, as they also focused on the reproduction of definitions, discussion of constituting elements in a concept, or explaining the reasons for a certain phenomenon. They did not seem to demand any kind of originality of expression, reasoning or thoughtful arguments.

Although the CS test included similar multiple choice and short answer test items, it differed from the IT and IBA tests in using inference and long answer test tasks that seemed to demand additional language skills. The test tasks on making inferences required students to paraphrase and apply previously learned concepts to new contexts; this arguably might need a good command of language to be accomplished. Likewise, the CS long answer test task entailed writing 800 to1000 word essay of novel language that should be based on both known or memorised information and individual judgements and thoughts (see Table 14). This task required building an argument about the definition of "Intercultural Communication", linking this to a theoretical background and supporting it with examples. Clearly this type of test tasks required a very good command of the English language.

Table 15. IT, IBA and CS Test Tasks Types and Examples from Spring 2009 Final Tests

Course	Test Task Type	Example	Weighting
IT	Multiple choice Question	_____ is one of the Arguments for Telecommuting.	30
	Short answer Question	List four benefits of E-commerce to society.	35
	Long answer Question	Briefly explain what is meant by "a system" and give three examples of systems.	15
Total			**80**
IBA	True/False Question	In the Hygiene (Two-Factor) Theory, workers work hard because they expect rewards for a good performance.	10
	Multiple choice Question	Franchised business in Oman is growing. Which is a *franchised* company?	10
	Short answer question	Define _culture_ and discuss _three (3)_ reasons why understanding it is important in business.	10
	Long answer Question	*List* and *explain* the _five (5)_ human needs according to Maslow's Hierarchy of Needs Theory.	20
Total			**50**
CS	inferring from a Reading test	Locate your example by indicating the line numbers and then paraphrase what is being communicated in these lines.	45
	Short answer question	What is conflict? Provide a definition of conflict. Please refer to communication terms (theory) in your answer.	30
	Long answer Question	Intercultural communication refers to communication between people who have different cultural beliefs, values or ways of behaving. Discuss this statement with reference to intercultural communication theory and give specific examples to illustrate these concepts. (800-1000 words essay)	25
Total			**100**

From the previous analysis of the academic course specifications, assessment schemes and test papers of the IT, IBA and CS introductory courses, it seems that the CS specialisation required a good command of the English language to successfully complete its assessment tasks much more than did the IT and IBA specialisations. Though from the CS course specifications alone this conclusion cannot be decisively made, the types of assessment instruments and test tasks used in the CS course revealed considerable demands on students' language skills. This finding could explain and consolidate the findings on FP the predictive validity; which indicated a strong correlation between students' scores in the FP assessment and CS assessment but a moderate correlation between their scores in the FP assessment and IT assessment and a non-significant correlation between FP and IBA assessment. This means that the predictive validity of FP assessment increases in the academic courses assessment when more command of the English language is required.

Discussion

This section is divided into three main subsections; the first deliberates on the general findings on FP predictive validity, while the second and third discuss the differences in predictive validity across specialisation groups and self-evaluation groups.

Predictive Validity of FP

Investigating the predictive validity of the FP English language assessment showed a significant but weak correlation between the students' grades in the FP English language assessment and their FY grade in academic courses. Students' grades in GES assessment showed a slightly stronger correlation coefficient with their grades in the academic courses assessment than did their grades in the AES assessment. This finding suggests that language proficiency does not predict academic achievement. This is in line with the conclusions drawn from similar previous studies conducted on the predictive validity of various English language tests that are used as gatekeepers to higher education institutions such as IELTS, TEAM, and various local tests (Davies, 1990; Elder, 1993; Cope, 2011; Lynch, 2000). Though these studies varied in the sample sizes, students' specialisations, levels of higher education, and measures of language proficiency and academic achievement, most of them concluded that the correlation between English language proficiency and academic achievement was weak to moderate, between 0.2 and 0.4.

This indicates that the predictive validity of FP assessment accounts only for about 16% of the variance of students' performance in academic courses assessment. However, it is sometimes stated that the correlation could be higher if the students who had not passed a pre-ssessional programme were included in predictive validity studies (Graham, 1987). Nonetheless, this finding stimulates questioning rigid policies on admitting students with specific language proficiency levels in higher education institutions not only in Oman but also in other international institutions. The difference in the strength of the predictive validity of GES and AES raises some questions about the reliability of performance assessment and consistency in using marking scales.

Predictive Validity of FP across Specialisations

This study reported that the strength of the correlation between the students' language proficiency and academic achievement varied considerably depending on the students' specialisations. These different predictive validity values for the specialisations could be partly explained by the language demands of these courses as reflected in their stated learning outcomes, assessment instruments and test tasks: the CS assessment instruments and test tasks seemed to draw upon students' language skills more than did those of the IT or IBA assessment instruments. In CS students are required to write a 1000 word reports, write two persuasive essays in the final exam and conduct presentations, all of which require a certain level of English language comand that is less required by the assessment tasks in other specialisation.

This finding along with similar findings from previous studies (e.g., Lynch, 2000; Huong, 2001) formulate a pattern that suggests a variation in language skills requiremetns of academic disciplines not only in Omani higher education, but also in many other international higher education institutions.

One of the implications of this finding is varying the entry level to higher education institutions based on the language requirements of the academic discipline. Though many international higher education institutions request different levels of language proficiency for different academic decVplines, these reuiremnts are usually not based on predictive validity studies. In Oman, CAS the should increase the language entry level for the students who wish to study Communication Studies.

This suggestion can be applied in other institutions if supported by local studies that include larger groups of participants.

Predictive Validity of FP across Self-Evaluation Groups

The correlations between language proficiency and academic achievement seemed to differ according to the students' self-evaluations of their language abilities. The higher the students evaluated themselves, the stronger the correlation between their grades in FP assessment and academic courses assessment became. Very few studies investigated the role self-evaluation played in academic achievement/difficulties, but the ones that did reported self-evaluation as a good predictor of academic difficulties (Xu, 1991). This suggests that more weight should be given to self-evaluation in future research on predictive validity. Also, self-evaluations can be applied in educational institutions as preliminary tool to understand students' academic achievement and/or difficulties.

Conclusion

In this paper, the predictive validity of the Foundation Programme assessment was explored by correlating students' scores in its assessment with their scores in the First Year academic courses. The findings revealed that proficiency in English is a moderate predictor of academic achievement in general. However the strength of the predictive validity was found to vary according to students' self-evaluations and specialisations, but not according to their gender or college. The higher the students evaluated their language proficiency, the higher the FP assessment predictive validity became. The predictive validity of FP assessment was strong for the CS and English language groups, moderate for the IT group and non-significant for the IBA group.

The findings of this study can feed into national educational policies in three ways. First, previous related findings (i.e., Author, 2013; Author, 2014) suggested that both students and teachers seemed to recognise that student language proficiency had a major impact in terms of accessing the labour market and higher education. With such a high-stakes assessment, its validity should be taken seriously to ensure that assessment uses and interpretations are supported by theoretical rationales and empirical evidence. Decisions linked with youth higher education opportunities or job opportunities are very critical and should be based on valid information.

The findings of this study reveal moderate to low predictive validity of English language assessment with regards to academic achievement, but students' proficiency in the English language plays a major role in accessing Omani higher education. Considering the findings of this study and other comparable ones, it is recommended that in admission to higher education, proficiency in English language should be considered as a criterion along with students' academic achievement, but used differently. Currently, higher education programmes that use English as a medium of instruction require a certain level of achievement in high school English language courses equal to that required in academic courses. Instead, if a high school graduate obtained the academic achievement level required, but not the English language level, he still should be considered for higher education admission but not if he meets the language requirement, but not the academic one.

Second, the AES assessment showed a lower value of its predictive validity than did the GES tests. This finding should be utilised in borderline cases where students' scores are very close to the cut-off point (50 out of 100). The present practice is that if a students' score is 48 or 49, it is rounded up to 50 (i.e., the passing score). I recommend that in such cases, students' scores in the GES assessment should be given more weight. This recommendation also supports the current policy followed in the FP of allowing admitted students to take a challenge exam (i.e., an English test offered to those who do well on the placement test which, if they pass will permit them to undertake FY courses without undergoing performance assessment tasks or taking FP English language courses).

English language assessment plays a critical role in Omani and international higher education and its impact is evident in higher education admission policies. This study suggests that these policies should be reviewed and interpretations made of student scores in English language assessment should be carefully considered in light of the findings of the predictive validity of FP assessment.

References

Author. (2014).

Author. (2013).

Al-Lamki, S. M. (1998). Barriers to Omanization in the private sector: The perceptions of Omani graduates. The International Journal of Human Resources Management, 9(2), 378-400.

Al-Musawi, N. M., & Al-Ansari, S. H. (1999). Test of English as a Foreign Language and First Certificate of English Tests as Predictors of academic sucess for undergraduate students at the University of Bahrain. System, 27, 389-399.

Alderson, J. C., Clapham, C., & Wall, D. (1995). Language Test Construction and Evaluation . Cambridge: Cambridge University Press.

Altbach, P. G., & Knight, J. (2007). The internationalization of higher education: Motivaion and realities. Journal of Studies in International Education, 11(3), 290-305.

Alsarimi, A. M. (2001). New trends in assessment in the Sultanate of Oman: Goals and characteristics. Educational Measurement: Issues and Practice, 27-29.

Bachman, L. F. (2004). Statistical analyses for language assessment. Cambridge: Cambridge University Press.

Bachman, L., & Plamer, A. (1996). Language testing in practice: Designing and developing useful language tests. Oxford: Oxford University Press.

Baker, D. (1989). Language testing: A critical survey and practical guide. London: Edward Arnold.

Breeze , R., & Miller, P. (2008). Predictive validity of the IELTS listening test as an indicator of student coping ability in Spain. IELTS Research Reports. 12, 1-34.

Bryman, A. (2008). Advances in mixed-methods research: Theories and applications. London: Sage Publications.

Davies, A. (1990). Principles of Language Testing. Oxford: Blackwell.

Dörnyei, Z. (2007). Reesrach Methods in Applied Linguistics and Mixed Methodologies. Oxford: Oxford University Press .

CAS. (2010). English Department Assessment Handbook. Muscat: Ministry of Higher Education.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT scores to academic performnace: Some evidence from American Universities. Language Testing, 29(3), 421-442.

Cope, N. (2011). Evaluating locally-developed language testing. Australian Review of Applied Linguistics, 34(1), 40-58.

Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongest a group of international students studying at the University of Tasmania. IELTS Research Reports, 1, 72-115.

Davies, A. (1990). Principles of Language Testing. Oxford: Blackwell.

Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. Melbourne Papers in Langauge Testing, 2(1), 68-89.

Feast, V. (2002). The impact of IELTS scores on performance at University. International Education Journal, 3(4), 70-85.

Fox, J. (2004). Test decisions over time: Tracking validity. Language Testing, 21(4), 437-465.

Graham, J. J. (1987). English language proficiency and the prediction of academic success. TESOL Quarterly, 21(3), 505-521.

Harrison, A. (1983). A language Testing Handbook. London: Mcmillan.

Hughes, A. (2003).Testing English for Language Teachers. Cambridge. UK: Cambridge University Press.

Huong, T. (2001). The predictive validity of the international English language test system (IELTS). The Post Graduate Journal of Education Research, 2(1), 66-96.

Jochems , W. (1991). Effects of learning and teaching in a foreign language. European Journal of Engineering Education, 16(4), 309-316.

Jochems, W., Sinppe, J., Jan Smid, H., & Verweij, A. (1996). The academic progress of foreign students: Study achievement and study behaviour. Higher Education, 31(3), 325-340.

Kane, M. (2011). Validating score interpretations and uses: Messick lecture. A paper presented in Langauge Testing Research Colloquium, (pp.3-17). UK: Cambridg..

Kerstjen, M., & Nery, C. (2000). Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance. IELTS Research Reports, 86-108.

Lynch, T. (2000). An evaluation of the revised test of English at matriculation at the University of Edinburgh. Edinburgh Working Papers in Applied Linguistics, 10, 61-71.

Maleki, A., & Zanjani, E. (2007). A survey on the relationship between English Language proficiency and the academic achievement of Iranian EFL students. Asian EFL Journal, 9(1), 86-96.

Martuza, V. R. (1977). Applying Norm-Referenced and Criterion-Referenced Measurement in Education. Boston: Allyan and Bacon.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. Educational Researcher, 18(5), 5-11.

Messick, S. (1995). Standards of Validity and the Validity of Standards in Performance Assessment. Educational Measurement. 14 (4). 5-8.

McNamara, T. F. (1996). Measuring second language performance. London and New York: Addison Wesley Longman.

Ministry of Higher Education. (2010). Higher Education Admission Statistics for the Academic Year 2009/2010. Muscat: Al-Karamel International for Media Services.

Oman Academic Accreditation Authority. (2009). Oman academic standards for general foundation programs. Muscat: OAAA.

Pallant, J. (2007). SPSS survival manual. New York: Open University Press.

Powers , D. E., Kim, H. J., & Weng, V. Z. (2008). The redesigned TOEIC (lisetning and reading) test: Relations to test-taker perceptions of proficiency in english. ETS Research Report.

Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. Language Testing, 25(1), 5-13.

Vinke, A., & Jochems, w. (1993). English proficiency and academic sucess in international postgraduate eductaion. Higher Education, 26, 275-285.

Weir, C. (2005). Language testing and validation. An evidence-based approach.UK: Palgrave MacMillan.

Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. University of Sydney Papers in TESOL, 1, 51-70.

Xu, M. (1991). The impact of English langauge proficiency on international graduate students' perceived academic difficulty. Research in Higher Education, 32(5), 557-569.

Yen, D. A., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. Worcester Journal of Learning and Teaching, 1-7.

Zabihi, R. (2011). Personality in English proficeincy and achievement. Continental Journal of Education Research, 4(1), pp.1-6